

**MIGUEL FILIPE GUILHERMINO DE MARINHO**

**MODELOS DE PREVISÃO  
PARA O PAR EUR/USD**



**DEPARTAMENTO DE MATEMÁTICA**

**FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO**

**2016**



MIGUEL FILIPE GUILHERMINO DE MARINHO

# MODELOS DE PREVISÃO PARA O PAR EUR/USD



*Tese submetida à Faculdade de Ciências da  
Universidade do Porto para obtenção do grau de Mestre  
em Engenharia Matemática*

ORIENTADOR: PROF. DOUTOR JOAQUIM FERNANDO PINTO DA COSTA

DEPARTAMENTO DE MATEMÁTICA  
FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO

2016



# Resumo

A previsão financeira é um tema de grande interesse para as comunidades económica e académica. Ter a capacidade de prever antecipadamente os movimentos do mercado cambial tem benefícios financeiros tanto para empresas como para investidores. Esta tese tem como objectivo dar a conhecer uma série de ferramentas matemáticas que podem ajudar a prever o mercado cambial de uma forma mais precisa.

A escolha do conjunto de teste tem influência no desempenho dum modelo, porém, não há um consenso de como dividir o conjunto de dados em treino, validação e teste. Assim, este trabalho engloba não só uma revisão às diferentes formas de divisão mais usadas na comunidade académica e económica, como também propostas de novas maneiras de o fazer.

Além disso, são aplicados modelos de regressão logística, árvores de decisão, análise discriminante linear e análise discriminante quadrática às diferentes divisões de dados sendo apresentados os diferentes resultados.

Por último são expostas três estratégias para dados do mercado cambial de alta frequência desenvolvidas no estágio curricular efectuado no Banco L. J. Carregosa.

**Palavras-chave:** PREVISÃO FINANCEIRA, MERCADO CAMBIAL, ANÁLISE TÉCNICA, MODELOS PREDITIVOS, METATRADER 4, EUR/USD.



# Abstract

Financial forecasting is a topic of much interest to both academic and economic communities. Being able to accurately forecast the foreign exchange rates is beneficial not only for firms but also for investors. This research aims to propose a set of mathematical tools that will enable to better anticipate the forex.

Out-of-sample tests of financial forecast performance depend on how a given data set is split into train, validation and test. However, no guidance exists on how to properly split the data set. So in this research, we not only reviewed the most popular ways to split the data set, but also proposed a few new ways of doing it.

In addition, we applied linear discriminant analysis, quadratic discriminant analysis, decision trees and logistic regression models to the different datasets and compared the results

Lastly, three MetaTrader 4's expertise advisors, which were developed during the internship I served at Banco L. J. Carregosa, are presented.

**Keywords:** FINANCIAL FORECASTING, FOREX, TECHNICAL ANALYSIS, PREDICTIVE MODELS, METATRADER 4, EUR/USD.





# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>3</b>
1.1	Problema . . . . .	4
1.2	Objectivo do Trabalho . . . . .	5
<b>2</b>	<b>Previsão Financeira</b>	<b>7</b>
2.1	Mercado Financeiro . . . . .	7
2.2	Mercado Cambial . . . . .	7
2.2.1	Operações Cambiais . . . . .	9
2.3	Previsibilidade dos Mercados . . . . .	11
2.4	Técnicas Tradicionais usadas no mercado cambial . . . . .	11
2.4.1	Análise Fundamental . . . . .	12
2.4.2	Análise Técnica . . . . .	12
<b>3</b>	<b>Indicadores Económicos</b>	<b>13</b>
3.1	Médias Móveis . . . . .	13
3.2	Oscilador Estocástico . . . . .	15
3.3	“Moving Average Convergence Divergence” (MACD) . . . . .	16
3.4	Taxa Verdadeira Média (ATR) . . . . .	17
<b>4</b>	<b>Modelos aplicados a dados de baixa frequência</b>	<b>19</b>
4.1	Divisão dos dados . . . . .	20
4.1.1	Ordem Cronológica . . . . .	21
4.1.2	Ordem Aleatória . . . . .	23
4.1.3	Divisão Mensal . . . . .	25

4.1.4	Subida/Descida/Constante . . . . .	26
4.2	Regressão logística . . . . .	40
4.2.1	Odds e Odds Ratio . . . . .	40
4.2.2	Construção do Modelo . . . . .	42
4.2.3	Modelo logit . . . . .	43
4.2.4	Resultados . . . . .	44
4.3	Árvores de Decisão . . . . .	45
4.3.1	Introdução . . . . .	45
4.3.2	Poda . . . . .	45
4.3.3	Resultados . . . . .	46
4.4	Análise Discriminante . . . . .	47
4.4.1	Linear . . . . .	47
4.4.2	Quadrática . . . . .	48
<b>5</b>	<b>Modelos aplicados a dados de alta frequência</b>	<b>51</b>
5.1	Plataforma MetaTrader 4 . . . . .	51
5.1.1	Estratégia 1 . . . . .	52
5.1.2	Estratégia 2 . . . . .	54
5.1.3	Estratégia 3 . . . . .	55
<b>6</b>	<b>Conclusão</b>	<b>57</b>

# Lista de Figuras

3.1 Média Móvel Simples (fonte: MetaTrader 4) . . . . .	14
3.2 Média Móvel Exponencial (fonte: MetaTrader 4) . . . . .	14
3.3 Média Móvel Simples (vermelho) e Média Móvel Exponencial (azul) . . . . .	14
3.4 Oscilador Estocástico (fonte: MetaTrader 4) . . . . .	15
3.5 MACD (fonte: MetaTrader 4) . . . . .	16
3.6 ATR (fonte: MetaTrader 4) . . . . .	17
4.1 Preço de fecho diário do EUR/USD entre 11/01/2001 e 23/01/2015 . . . . .	20
4.2 Divisão em treino, validação e teste por ordem cronológica . . . . .	22
4.3 Divisão em treino, validação e teste aleatoriamente . . . . .	23
4.4 Divisão em treino, validação e teste por mês . . . . .	25
4.5 Divisão dos dados em períodos . . . . .	26
4.6 Período 1 . . . . .	27
4.7 Período 2 . . . . .	28
4.8 Período 3 . . . . .	29
4.9 Período 4 . . . . .	30
4.10 Período 5 . . . . .	31
4.11 Período 6 . . . . .	32
4.12 Período 7 . . . . .	33
4.13 Período 8 . . . . .	34
4.14 Período 9 . . . . .	35
4.15 Período 10 . . . . .	36
4.16 Divisão em treino, validação e teste por períodos . . . . .	37

4.17 Divisão do Período 1 em treino, validação e teste . . . . .	38
4.18 Divisão em treino, validação e teste por períodos . . . . .	39
4.19 Árvore obtida usando a divisão de dados aleatória . . . . .	46
4.20 Árvore obtida usando a divisão de dados por períodos 2 . . . . .	47



# 1 Introdução

Com um volume médio diário de  $5.3 \cdot 10^{11}$ \$ [1], o mercado cambial é de longe o maior mercado financeiro. Este mercado é estudado através de complexas séries temporais (não estacionárias) dos preços de câmbio (não lineares) [2]. Estas características combinadas com outros fatores económicos, políticos e sociais, tornam a previsão dos preços de câmbio uma das mais difíceis aplicações da área de “forecasting”.

A flutuação dos preços de câmbio tem um grande impacto na economia global, desde grandes corporações que vendem os seus produtos em várias moedas a instituições financeiras e pequenos investidores. Prever com precisão o mercado cambial é essencial à saúde financeira dos seus portfolios. Assim, a motivação deste trabalho é aplicar vários modelos de estatística e “machine learning” de forma a prever o mercado cambial com a maior precisão possível.

Esta tese foi escrita a par da realização dum estágio curricular no Banco L. J. Carregosa, onde me foi proposto a implementação dum sistema de negociação automático na plataforma MetaTrader 4. Assim ao longo da tese poderão ser encontrados resultados relativos ao estágio curricular, assim como resultados de investigação realizada na Faculdade de Ciências da Universidade do Porto sob orientação do Professor Doutor Joaquim Pinto da Costa.

### 1.1 Problema

A previsão financeira é um tópico de muito interesse tanto para a comunidade económica como académica [3]. Dentro da comunidade académica, a previsão do mercado cambial é um tópico muito tratado, geralmente através de várias técnicas de data mining. Algumas destas técnicas vão ser abordados ao longo desta tese como por exemplo: Regressão logística, Análise Discriminante Linear, Análise Discriminante Quadrática, Árvores de Decisão. Estes modelos que também são por vezes chamados de modelos de inteligência artificial, são aplicados com sucessos em muitas outras áreas como biologia, geologia, astronomia e medicina [4].

Dentro da comunidade económica a previsão do mercado tem benefícios a nível de ganhos financeiros diretos ou diminuição de risco. Por exemplo, a correta previsão e compreensão do mercado cambial pode levar à realização de acordos entre instituições com grande vantagem para ambas as partes como aconteceu com o contrato “Swap” entre o Banco Mundial e a International Business Machines [5]. Por outro lado, quando não é feita uma correta abordagem de todos os fatores financeiros intervenientes, este tipo de contratos pode originar grandes perdas para uma das partes, como por exemplo os contratos "swap", efetuados entre o banco Santander Totta e empresas públicas Portuguesas, têm perdas potenciais para as empresas públicas de  $1.5 \cdot 10^9 \text{€}$

## 1.2 Objectivo do Trabalho

O principal objetivo desta tese é propor vários modelos de previsão do mercado cambial, assim como explicar detalhadamente a construção dos mesmos, de forma a que estes possam servir de suporte tanto a instituições como a investidores que pretendam prever os preços de câmbio.

Relativamente ao trabalho levado a cabo no estágio curricular no Banco L. J. Carregosa, foi-me proposta a implementação dum sistema de negociação automático em ambiente MetaTrader 4 (ver capítulo 5). Esta instituição, que teve origem na casa financeira L. J. Carregosa (a mais antiga da Península Ibérica), criada em 1833 no Porto, lançou o primeiro serviço de corretagem online em Portugal. Foi-me dito que o sistema que implementasse deveria satisfazer as seguintes condições:

- Ter uma rentabilidade anual igual ao superior a 5% no ano de 2015
- Colocar uma média de 400 ordens (de compra ou venda) no mercado cambial diariamente
- Ter uma perda máxima de 20%

Durante a implementação deste sistema de negociação, foram surgindo questões relativamente a procedimentos usuais dentro das comunidades académica e económica relativamente a tratamento de dados, treino e testes de modelos. Estes problemas motivaram uma posterior pesquisa levada a cabo na Faculdade de Ciências da Universidade do Porto.

Na comunidade económica há dois tipos de previsão de mercado: análise técnica, que se baseia puramente em dados históricos como preço e volume, e a análise fundamental que tenta prever através de fatores macro e microeconómicos. Este trabalho foca-se exclusivamente na análise técnica, mais precisamente apenas utiliza o histórico de preços do EUR/USD, por que foi escolhido por ter o maior volume de ordens diário [1].





## 2 Previsão Financeira

### 2.1 Mercado Financeiro

O mercado financeiro permite a compra e venda de ativos financeiros a pessoas e instituições. As ações, que são um dos ativos mais populares, representam uma parte duma empresa cotada em bolsa. Outros ativos são as mercadorias, obrigações e moeda [7]. Todos estes ativos financeiros são diariamente comprados e vendidos nos seus respetivos mercados que estão devidamente regulados. A atenção deste trabalho recai sobre o mercado cambial, mais precisamente sobre o par EUR/USD.

### 2.2 Mercado Cambial

O mercado cambial é, sem dúvida, o maior mercado financeiro, com um volume diário superior a  $5.0 \cdot 10^{11}$  \$ [1], é a base para todo o comércio internacional. Em termos internacionais é conhecido por "Foreign Exchange Market", "Forex" ou "FX". Este mercado teve origem no início da década de setenta do século passado [8] e veio pôr fim ao regime cambial de taxas fixas (acordo de Bretton-Woods), iniciando, assim, um regime de câmbios flutuantes, em função da procura e da oferta de moeda.

As principais funções do mercado cambial são:[8]

- facilitar o encontro entre a oferta e a procura de moeda, de que resulta a definição do preço ou a cotação das diferentes divisas, isto é, a fixação das taxas de câmbio;
- fomentar as trocas comerciais a nível internacional, através de remessas e de créditos documentários, peças essenciais para a concretização de exportações e de importações;

- possibilitar a liquidação financeira de transacções entre pessoas ou entidades de diferentes países (transacções comerciais e operações financeiras);
- permitir a gestão dos riscos cambiais, com base em operações cambiais a prazo, tais como "forwards" cambiais (contrato financeiro que fixa a taxa de câmbio para uma operação financeira a realizar no futuro), "swaps" de moeda, opções e contratos futuros.

Atualmente, as moedas com um maior volume de ordens no mercado cambial são o dólar americano (USD), o euro (EUR), o yen japonês (JPY), a libra esterlina (GBP), o dólar australiano (AUD) e o franco suíço (CHF). [1] As principais características do mercado cambial são as seguintes:[8]

- é um mercado global em que a formação dos preços só depende da procura e da oferta das diferentes moedas num determinado momento;
- é um mercado OTC - "Over the Counter" ou mercado de balcão, em que cada transacção é conduzida e controlada por ambas as partes e em que definem as condições da mesma, no que diz respeito a preço, valor da transacção e data da liquidação financeira. As transacções têm por base a informação disponibilizada por adequados sistemas de informação, como a Bloomberg, a Reuters ou a Bridge. As principais praças financeiras do mercado cambial são Londres e Nova Iorque;
- o mercado cambial funciona 24 horas por dia;
- é um mercado eficiente, uma vez que os preços do mercado reflectem toda a informação disponível considerada relevante;
- é um mercado com elevada liquidez, o que permite abrir e fechar posições em questão de segundos.

Os bancos centrais, os bancos comerciais, os bancos de investimento, os corretores de câmbios e os clientes são os principais intervenientes do mercado "FX". As ordens de compra e venda de moeda são dadas pelos clientes aos bancos comerciais e de investimento, que por sua vez estão em contacto com corretoras de câmbios ou outras entidades especializadas,

responsáveis pela concretização do negócio. Por fim, a liquidação financeira das operações é realizada ao nível do sistema financeiro, no mercado monetário interbancário. Todo este processo é supervisionado pelo Banco Central.

### 2.2.1 Operações Cambiais

Uma operação cambial consiste num acordo entre duas entidades, no qual é fixada uma data e um preço pelo qual ambas as partes se comprometem a trocar uma determinada quantidade de moeda por outra.

Os elementos de uma operação cambial são os seguintes: [8]

- duas moedas diferentes, o que implica uma compra e uma venda simultânea;
- a taxa de câmbio que é o valor de uma moeda expressa em unidades de outra moeda;
- o prazo da operação, que é o número de dias que decorrem entre a data do acordo e a data de entrega efectiva das moedas da operação (data de contratação e data da liquidação).

A taxa de câmbio, também conhecida por câmbio ou paridade cambial, é definida no mercado, em função das condições e das informações disponíveis a cada momento, e representa o preço relativo entre duas moedas. Esta taxa é sempre apresentada por uma abreviatura, as moedas são representadas por uma abreviatura de três letras, em que as duas primeiras representam o país e a última é a letra inicial do nome da moeda. [9]

A paridade cambial é a quantidade da segunda moeda que precisamos de dar para receber uma unidade da primeira. Se pensarmos no  $\text{EUR/USD} = 1.32$ , isto significa que 1 euro vale 1.32 dólares americanos. Neste caso, o EUR funcionou como valor certo (uma unidade) e o USD funcionou como um valor incerto (flutua com o mercado). Este sistema designa-se por cotação ao certo ou sistema de cotação indirecta. [8] Se  $1 \text{ EUR} = 1.32 \text{ USD}$ , então  $1 \text{ USD} = 0.7576 \text{ EUR}$ , neste caso diz-se que a cotação é ao incerto ou sistema de cotação directo.

O mercado cambial funciona com duas taxas de câmbio: o bid (preço de compra) e o ask (preço de venda), estas taxas de câmbio são apresentadas da seguinte forma:

EUR/USD 1.3200-1.3210

Isto significa que o mercado compra 1 EUR por 1.32 USD e vende 1 EUR por 1.3210 USD, o valor absoluto da diferença destes valores denomina-se margem comercial do banco ou "spread".

Existem algumas convenções no mercado Forex: [8]

- a cotação de compra é apresentada em primeiro lugar e a cotação de venda é apresentada em segundo lugar;
- a divisa de base é a moeda com maior valor absoluto, logo, é apresentada em primeiro lugar
- as casas decimais variam de moeda para moeda mas uma vez definidas, mantêm-se constantes para cada moeda. A forma de apresentação pode ser a seguinte: EUR/USD: 1.3200-10, o que significa 1.3200-1.3210;
- as cotações de compra e de venda são sempre apresentadas na óptica do mercado (intermediário financeiro). O banco compra e o banco vende. Quem deseja comprar uma moeda, compra ao preço que o mercado vende e quem deseja vender uma moeda, vende ao preço a que o mercado compra.

Existem dois tipos de operações cambiais:

- operações cambiais à vista ou "spot", que consistem na negociação e troca de moedas em simultâneo e com base numa taxa de câmbio que está a vigorar nesse dia, sendo que a liquidação financeira da operação ocorre normalmente passados dois dias úteis (a data-valor de cada operação corresponde ao segundo dia útil após a data de concretização da operação cambial);
- operações cambiais a prazo ou "forward", ou seja, neste caso o período entre o momento da negociação e o da troca efectiva das moedas é superior a 2 dias. A taxa de câmbio a considerar é a que vigorar no dia da contratação da operação; contudo, a data-valor dista mais de 2 dias da data de negociação.

## 2.3 Previsibilidade dos Mercados

Há muitas aplicações que tentam prever o mercado financeiro. Porém, a verdadeira questão é se será realmente possível prever uma estrutura tão complexa como o mercado financeiro? Por exemplo, a “random walk hypothesis” é uma teoria financeira que propõe que os preços dos mercados financeiros evoluem segundo um passeio aleatório, o que implica que não são previsíveis. A teoria remonta ao ano de 1863, data da publicação do livro “Calcul des chances et philosophie de la bourse” por [10], corretor de bolsa francês que se tornou milionário. Posteriormente, o matemático francês Louis Bachelier na sua tese de doutoramento intitulada “Théorie de la spéculation” [11], também defendeu a imprevisibilidade do mercado financeiro. Mais recentemente outros trabalhos semelhantes foram publicados, com destaque para [12] e [13]. Fama concluí, baseado em teste empíricos, que o histórico de preços suporta a “random walk hypothesis”. Porém, a teoria desenvolvida por Fama apenas se baseia em modelos lineares para os mercados financeiros [14]. Contudo a não dependência linear dos mercados, não descarta a existência de dependências não lineares o que contradiria a “random walk hypothesis”. De facto, estudos recentes vêm apoiar a existência de dependências não lineares e por isso a possibilidade de previsão dos mercados [15]. Assim, visto que existe a possibilidade de o mercado financeiro ser de facto previsível, surge naturalmente o interesse em tentar prever o mesmo com a maior exatidão possível. Por último, gostaria de mencionar que conseguir prever o mercado com sucesso, não se traduz automaticamente em ganhos financeiros. Para efetivamente tornar a atividade lucrativa, terá que se ter em consideração todos os custos inerentes ao mercado (como por exemplo: comissões de corretagem, tributações, etc).

## 2.4 Técnicas Tradicionais usadas no mercado cambial

As técnicas tradicionalmente usadas no mercado cambial dividem-se em dois tipos de análise: [16] análise técnica e análise fundamental. Ambos os tipos de análise foram muito populares no passado e poderão de alguma forma dar um contributo positivo à investigação. Obviamente, será dada uma maior atenção à análise técnica por estar mais relacionada com

o trabalho de investigação desenvolvido.

### 2.4.1 Análise Fundamental

A análise fundamental baseia-se em fatores macro e/ou microeconómicos que influenciam o preço dum dado ativo financeiro para tentar prever o seu valor futuro. No caso do mercado cambial são maioritariamente usados fatores macroeconómicos, como por exemplo: taxas de juro, taxas de inflação e a taxa de desemprego para tentar prever os preços de câmbio duma dada moeda. Porém, existe controvérsia na previsão financeira através da análise fundamental. Em 1983, Meese e Rogoff [17] mostram que prever o mercado cambial através do passeio aleatório tem um melhor desempenho do que prever através de modelos microeconómicos. Mais recentemente, a investigação de Meese e Rogoff é revista e confirmada por [18]. Contudo muitos investigadores defendem a existência de relação entre preços de câmbio e alguns acontecimentos macroeconómicos.

### 2.4.2 Análise Técnica

A análise técnica baseia-se na informação do mercado do passado, para tentar prever o futuro. A informação usada, normalmente, resume-se ao histórico de preços, porém outras informações podem ser usadas, como por exemplo o volume. O volume consiste na soma do valor de todas as ordens colocadas no mercado num dado período temporal. A análise técnica é altamente popular no mercado cambial e são geralmente usados um conjunto de indicadores económicos que passamos a descrever.

## 3 Indicadores Económicos

Vamos tomar:

- $price(t)$  - o preço de câmbio no instante  $t$ .
- $close(t)$  - o último preço de fecho anterior ao instante  $t$
- $high(t)$  - o máximo dos preços do dia anterior ao dia do instante  $t$
- $low(t)$  - o mínimo dos preços do dia anterior ao dia do instante  $t$

### 3.1 Médias Móveis

As duas variantes mais populares deste indicador são: Média Móvel Simples (“Simple Moving Average” (SMA)) e Média Móvel Exponencial (“Exponential Moving Average” (EMA)). Apesar de já não serem muito utilizados diretamente, são a base de muitos outros indicadores (exemplo: MACD, “Bollinger Bands”, etc). A diferença entre os dois tipos está no facto do EMA dar mais peso aos preços mais recentes, enquanto que os SMA é uma simples média aritmética.

Cálculo [19]:

$$SMA(price(t), n) = \frac{price(t) + price(t-1) + \dots + price(t-(n-1))}{n} \quad (3.1)$$

$$EMA(price(t), n) = EMA[price(t-1, n)] + [close(t) - EMA(price(t-1), n)] * \frac{2}{n+1} \quad (3.2)$$





**Figura 3.1** Média Móvel Simples (fonte: MetaTrader 4)



**Figura 3.2** Média Móvel Exponencial (fonte: MetaTrader 4)



**Figura 3.3** Média Móvel Simples (vermelho) e Média Móvel Exponencial (azul)

## 3.2 Oscilador Estocástico

Desenvolvido no final do anos 50, este indicador compara o preço atual duma mercadoria com o preço que a mesma mercadoria custou nos últimos  $k1$  instantes e é constituído por duas curvas denominadas  $K(t)$  (linha vermelha Figura 3.4) e  $D(t)$  (linha azul Figura 3.4). Geralmente, quando  $D(t)$  se encontra acima de 80 diz-se que a mercadoria está “overbought” por outro lado quando está abaixo de 20 classifica-se como “oversold”. A curva  $K(t)$  é usada por alguns corretores para detetar sinais de compra ou venda quando esta se intersesta com o  $D(t)$ .

Cálculo: [20]

$$\min(\text{price}(t), k1) = \min[\text{price}(t-1), \text{price}(t-2), \dots, \text{price}(t-k1)] \quad (3.3)$$

$$\max(\text{price}(t), k1) = \max[\text{price}(t-1), \text{price}(t-2), \dots, \text{price}(t-k1)] \quad (3.4)$$

$$K(t) = \frac{\text{price}(t) - \min(\text{price}(t), k1)}{\max(\text{price}(t), k1) - \min(\text{price}(t), k1)} \quad (3.5)$$

$$D(t) = \text{SMA}(K(t), k2) \quad (3.6)$$



**Figura 3.4** Oscilador Estocástico (fonte: MetaTrader 4)

### 3.3 “Moving Average Convergence Divergence” (MACD)

Desenvolvido no final do anos 70, este indicador é um dos mais populares e simples. Baseia-se na diferença de duas EMA com diferentes períodos, podendo, assim, tomar valores positivos e negativos. É, assim, constituído por duas linhas chamadas Main (barras cinzentas Figura 3.5) e Signal (linha vermelha Figura 3.5).

Cálculo:

- K1 – período da 1ª Média Móvel Exponencial (valor standard é 12)
- K2 – período da 2ª Média Móvel Exponencial ( $k_2 > k_1$ ) (valor standard é 26)
- K3 - período da Média Móvel Simples (valor standard é 9)

$$Main = EMA(price(t), K1) - EMA(price(t), K2) \quad (3.7)$$

$$Signal = SMA(price(t), K3) \quad (3.8)$$



**Figura 3.5** MACD (fonte: MetaTrader 4)

### 3.4 Taxa Verdadeira Média (ATR)

A Taxa Verdadeira Média (“Average True Rate”) é uma medida de volatilidade introduzida por Welles Wilder no seu livro: *New Concepts in Technical Trading Systems*. [21] O ATR aumenta proporcionalmente com a volatilidade do  $price(t)$ . Seja:

- $high(t^*)$  - o máximo dos preços do dia verificado até ao instante  $t$
- $low(t^*)$  - o mínimo dos preços do dia verificado até ao instante  $t$

Cálculo: [22]

$$TR(t) = \max[high(t^*) - low(t^*), |high(t^*) - close(t)|, |low(t^*) - close(t)|] \quad (3.9)$$

$$ATR(t) = EMA(TR(t), C1) \quad (3.10)$$



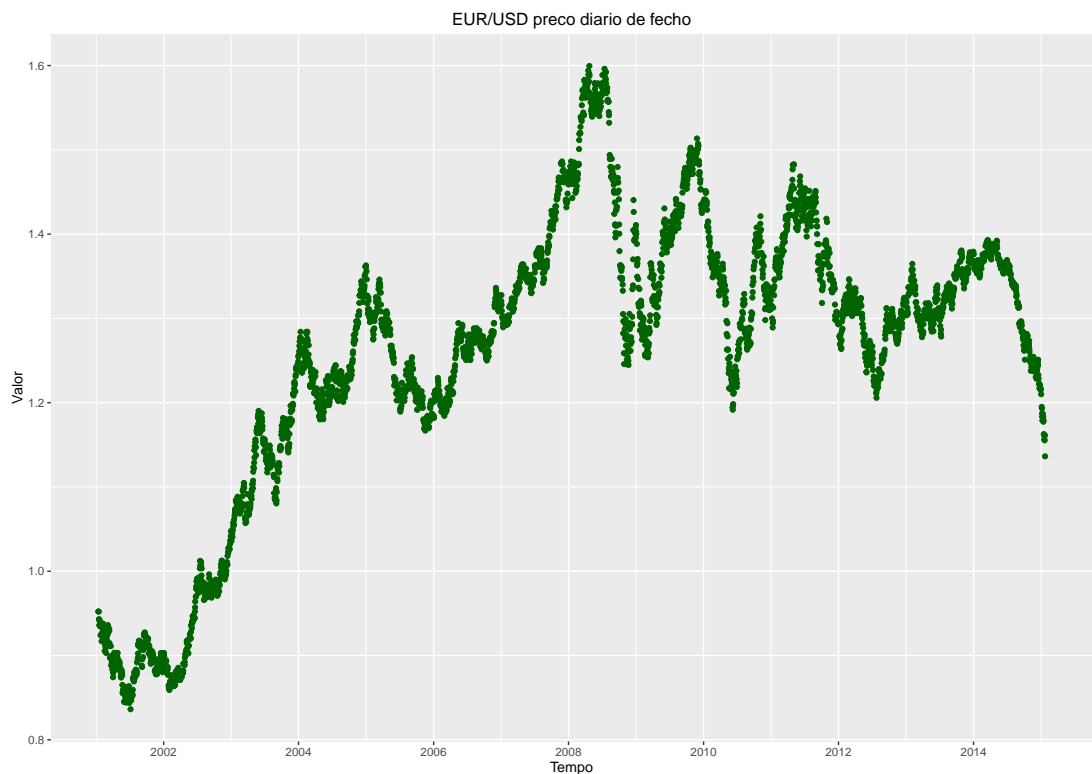
**Figura 3.6** ATR (fonte: MetaTrader 4)



## 4 Modelos aplicados a dados de baixa frequência

Nesta tese são usados dados retirados do Yahoo Finance, mais precisamente foram usados os preços de fecho diários do EUR/USD de 11/01/2001 a 23/01/2015 num total de 3648 dias. Foram, também, criadas um conjunto de variáveis:

- Date: Data do preço de fecho (no dia t)
- Close:  $\text{Close}(t)$
- Lag1:  $\text{Close}(t-1)$
- Lag2:  $\text{Close}(t-2)$
- ...
- Lag7:  $\text{Close}(t-7)$
- Min:  $\min[\text{Lag1}, \text{Lag2}, \dots, \text{Lag7}]$
- Max:  $\max[\text{Lag1}, \text{Lag2}, \dots, \text{Lag7}]$
- K:  $100 * \frac{\text{Lag1} - \text{Min}}{\text{Max} - \text{Min}}$
- D: média aritmética dos 3 últimos K
- Direction: Down se  $\text{Close}(t) < \text{Lag1}$ , Up caso contrário



**Figura 4.1** Preço de fecho diário do EUR/USD entre 11/01/2001 e 23/01/2015

### 4.1 Divisão dos dados

Quando o nosso objetivo é prever, mais especificamente desenvolver modelos de previsão, normalmente vamos utilizar estes modelos para suportar decisões. Quando utilizamos um modelo preditivo o nosso foco já não está nos dados passados, mas em utilizar esses dados para tentar criar uma teoria sobre a realidade, que nos permita de alguma forma prever o futuro. Em “terreno desconhecido” temos todas as razões para sermos céticos sobre os resultados que o modelo fornece, surgindo, assim, a necessidade de criar uma forma de avaliar a precisão e consistência do nosso modelo.

A forma mais simples de avaliar o desempenho do modelo no futuro é simulá-lo de alguma forma. Obviamente, nunca vamos ter acesso ao futuro, porém, podemos reservar parte dos dados que temos ao nosso dispor e tratá-los como se fossem dados do futuro. Por exemplo,

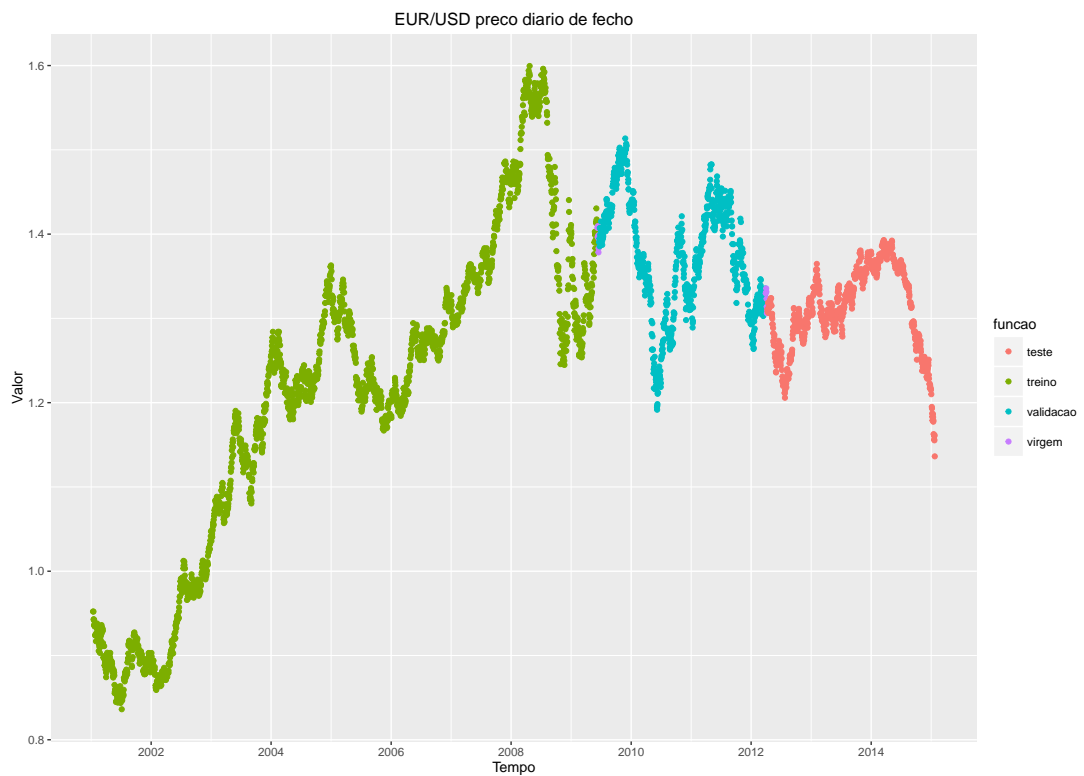
na previsão de preços de câmbio, podemos usar os dados disponíveis de 2014 para construir o modelo e fazer previsões sobre os preços de 2015. Visto que também temos acesso aos preços de 2015, podemos comparar as nossas previsões com a realidade de 2015, tendo assim uma ideia do desempenho do nosso modelo.

A partição mais simples e mais utilizada em previsão financeira consiste em utilizar os dados mais recentes para teste, geralmente chamada “out-of-sample analysis” e os restantes para treino. Neste capítulo pretende-se que sejam apresentadas várias formas de partição do conjunto de dados.

##### 4.1.1 Ordem Cronológica

Como foi referido anteriormente, a partição do conjunto de dados associada à ordem cronológica é a mais usada na comunidade; os dados mais recentes são usados para teste e os restantes para treino. Sendo este o método usual, não podíamos deixar de fora desta tese. Porém, visto que pretendemos fazer escolhas entre modelos, decidimos fazer uma partição extra, obtendo um novo conjunto a que chamamos de validação, tal como é sugerido por [23].



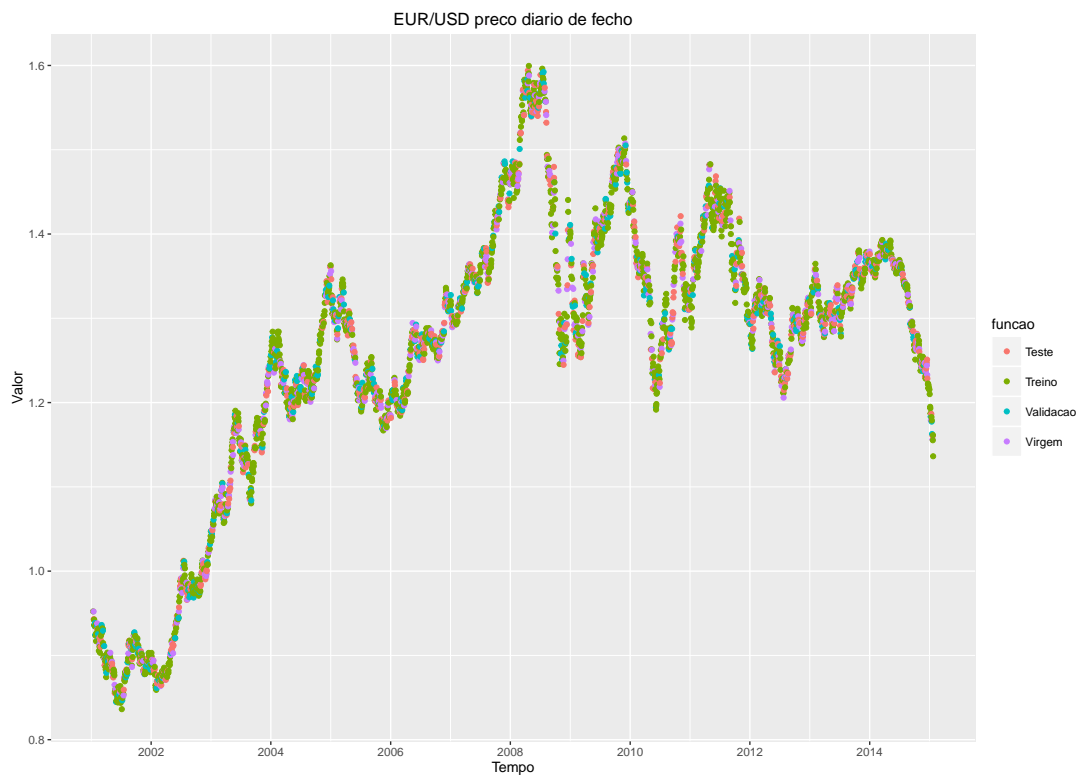


**Figura 4.2** Divisão em treino, validação e teste por ordem cronológica

Esta não parece ser a melhor forma de treinar, validar e testar um modelo. Vamos supor que desde o dia 2500 o EUR/USD começa a ter um determinado comportamento nunca antes visto. Como o modelo não foi treinado com esse novo comportamento pode não estar treinado da melhor forma.

### 4.1.2 Ordem Aleatória

Uma alternativa seria dividir os dados em treino, teste e validação de forma aleatória. Porém, visto que a nossa observação do  $i$ -ésimo dia contém informação dos 7 dias anteriores, ao dividir aleatoriamente os dados vai ser muito provável que, por exemplo, o treino tenha informação sobre o teste, isto é, alguns dos “lags” das observações do treino vão ser preços de fecho de teste. Uma forma de solucionar o problema, seria de cada observação que era aleatoriamente escolhida para treino, teste ou validação remover as 7 anteriores de forma a garantir que não há “cruzamento” de informação entre os conjuntos. Porém, não temos observações suficientes para aplicar tal método. Assim, a alternativa que se segue não resolve o problema do cruzamento de informação por completo. O que se propõe é que o teste seja escolhido de forma aleatória, as observações do dia anterior a um dia de teste sejam descartadas e que as observações restantes sejam aleatoriamente divididas em treino e validação.

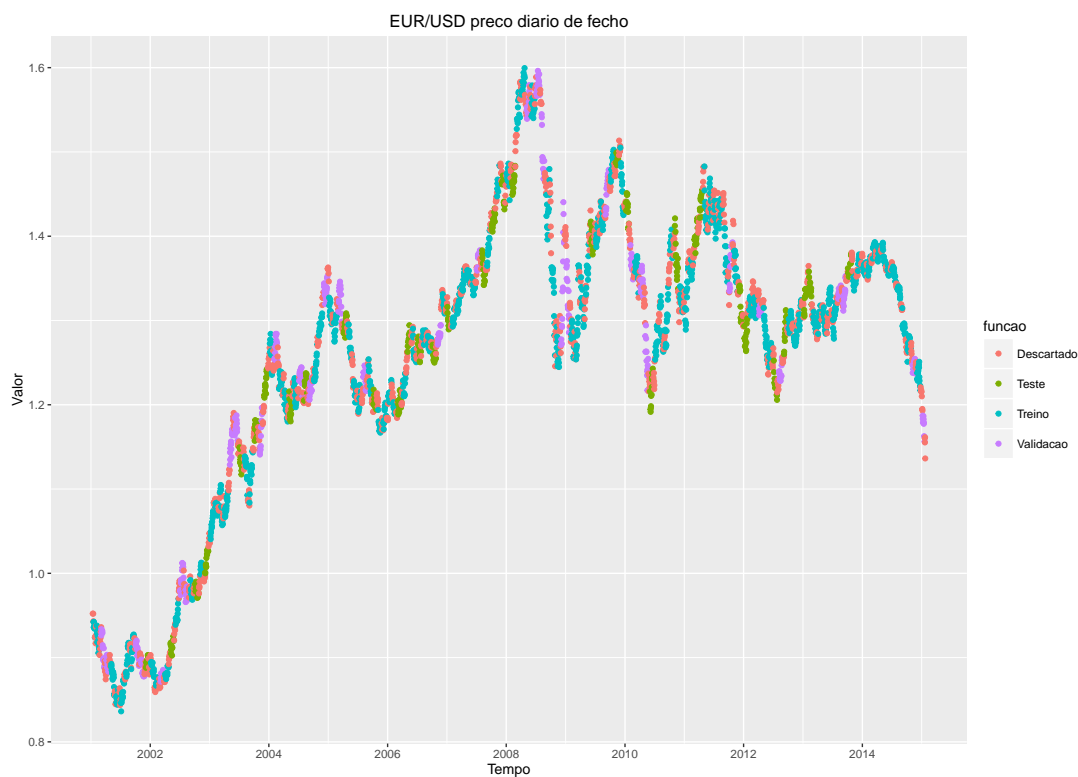


**Figura 4.3** Divisão em treino, validação e teste aleatoriamente

Obviamente, esta alternativa está longe de ser ideal, visto que há demasiado cruzamento de informação entre os conjuntos, especialmente entre o treino e a validação. Por isso, decidimos criar outras alternativas para a divisão dos dados.

### 4.1.3 Divisão Mensal

Para evitar o cruzamento de informação entre os três conjuntos, é necessário usar períodos contínuos de tempo para treino, validação e teste. Uma das formas mais naturais de o fazer é escolher meses aleatórios para cada um dos conjuntos tendo o cuidado de descartar observações, de forma a não haver cruzamento de informação entre conjuntos. Há várias formas possíveis de descartar as observações por exemplo remover as sete primeiras observações de cada mês seria talvez a mais natural porém, nesta tese decidimos remover as três primeiras e as quatro últimas observações de cada mês, garantindo, assim, que não há cruzamento de informação entre meses.

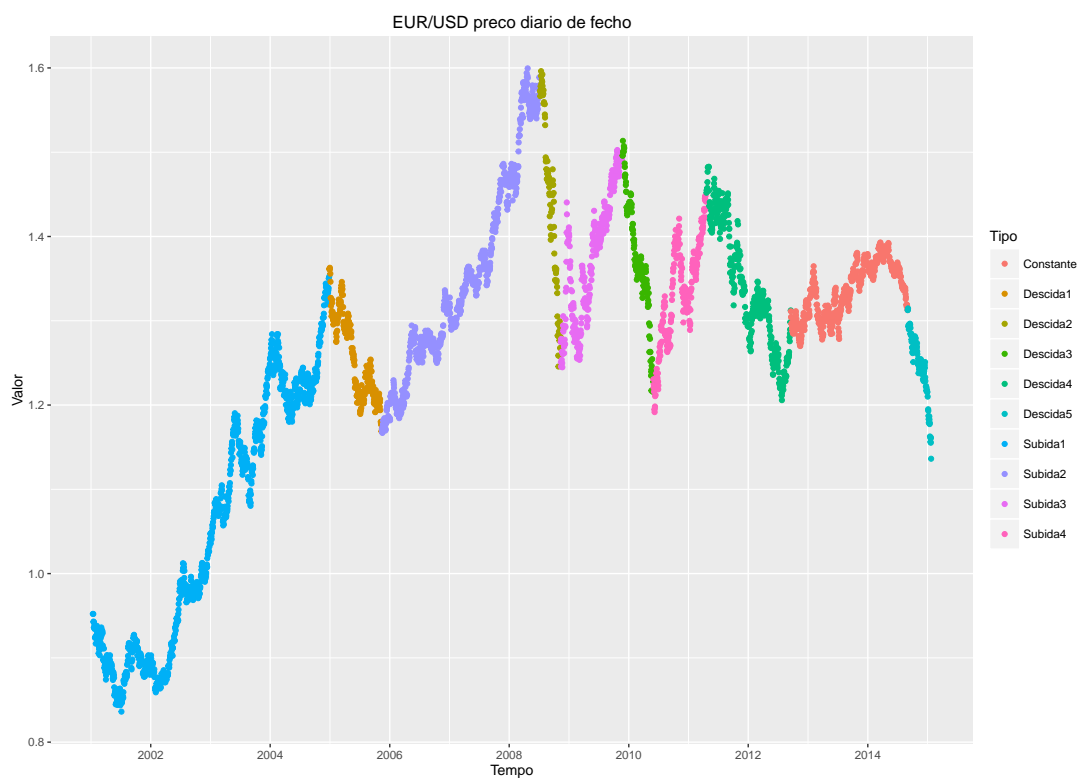


**Figura 4.4** Divisão em treino, validação e teste por mês

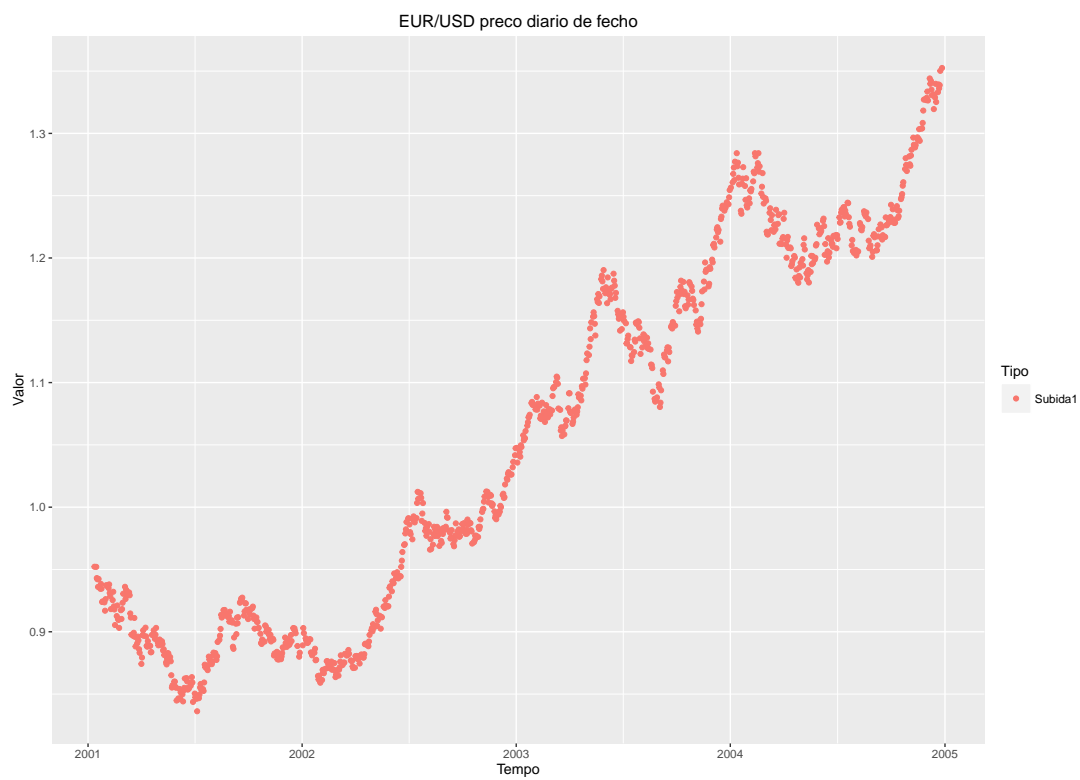
Ainda assim, à volta de 30% das observações são descartadas para evitar o cruzamento de informação, o que nos levou a formular outro método de divisão que nos permita não perder tantas observações e manter a “independência” entre conjuntos.

#### 4.1.4 Subida/Descida/Constante

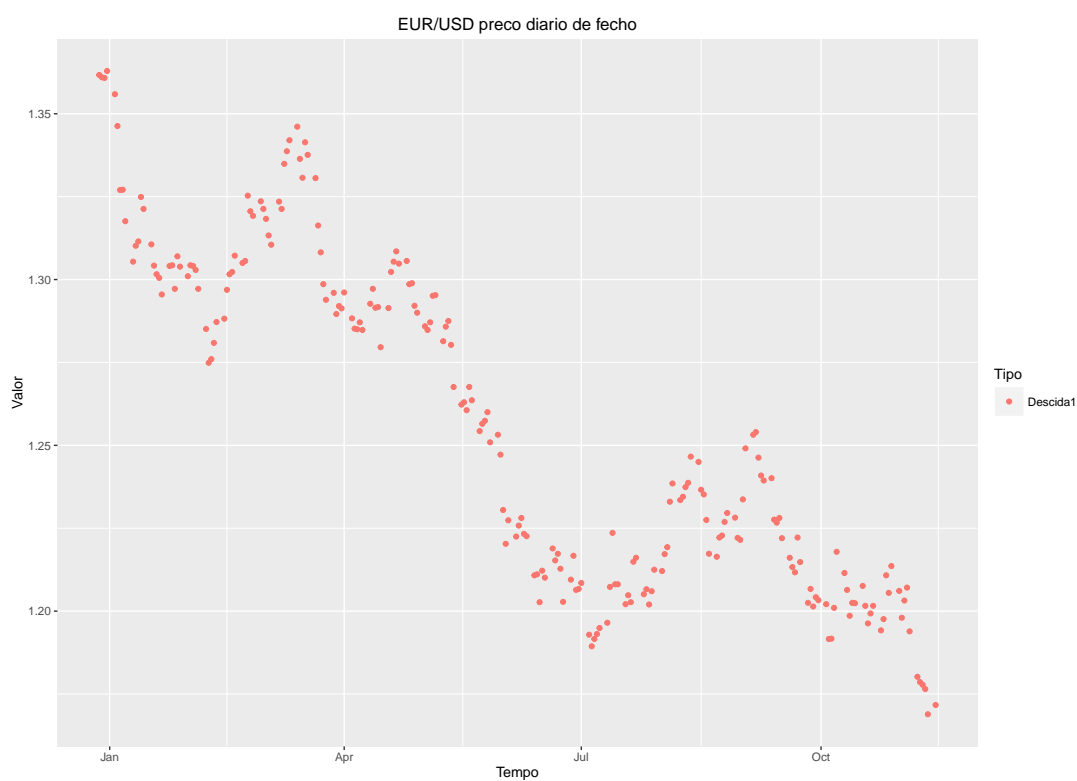
Obviamente, para conseguir o que pretendemos temos que aumentar os períodos contínuos de tempo uma maneira óbvia seria em vez de dividir por meses, dividir por anos. Porém, decidimos fazer de forma diferente e desta vez apresentamos a divisão com base em períodos por tendências (subida, descida ou constante). Vamos, assim, dividir o conjunto de dados em dez períodos, sendo cada um dos quais “avaliado” como “subida”, “descida” ou “constante” tendo em conta o aspeto visual do gráfico. As sete últimas observações de cada período serão descartadas para garantir que não há cruzamento de informação entre períodos.



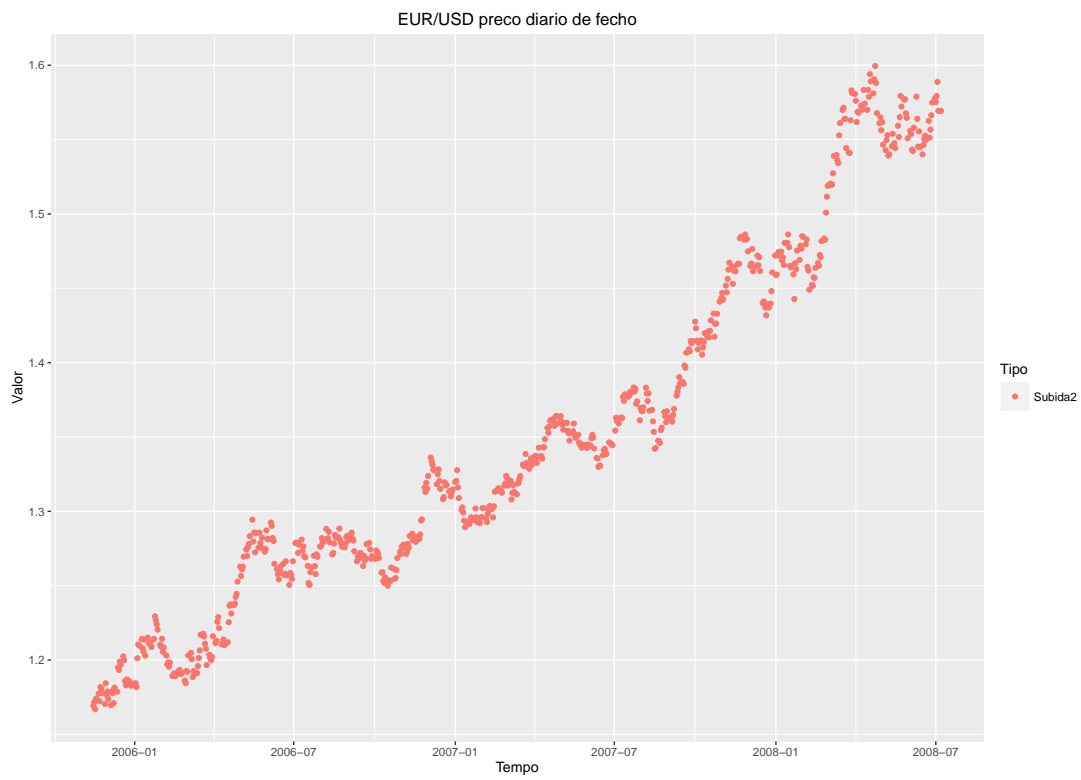
**Figura 4.5** Divisão dos dados em períodos



**Figura 4.6** Período 1

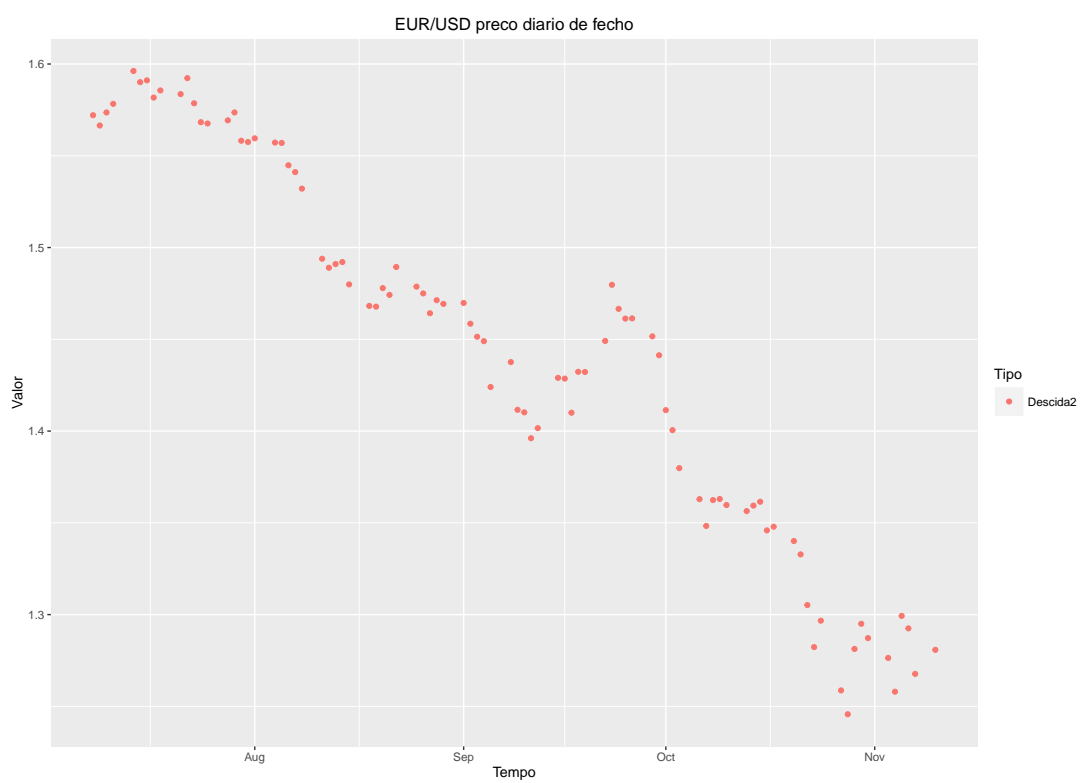


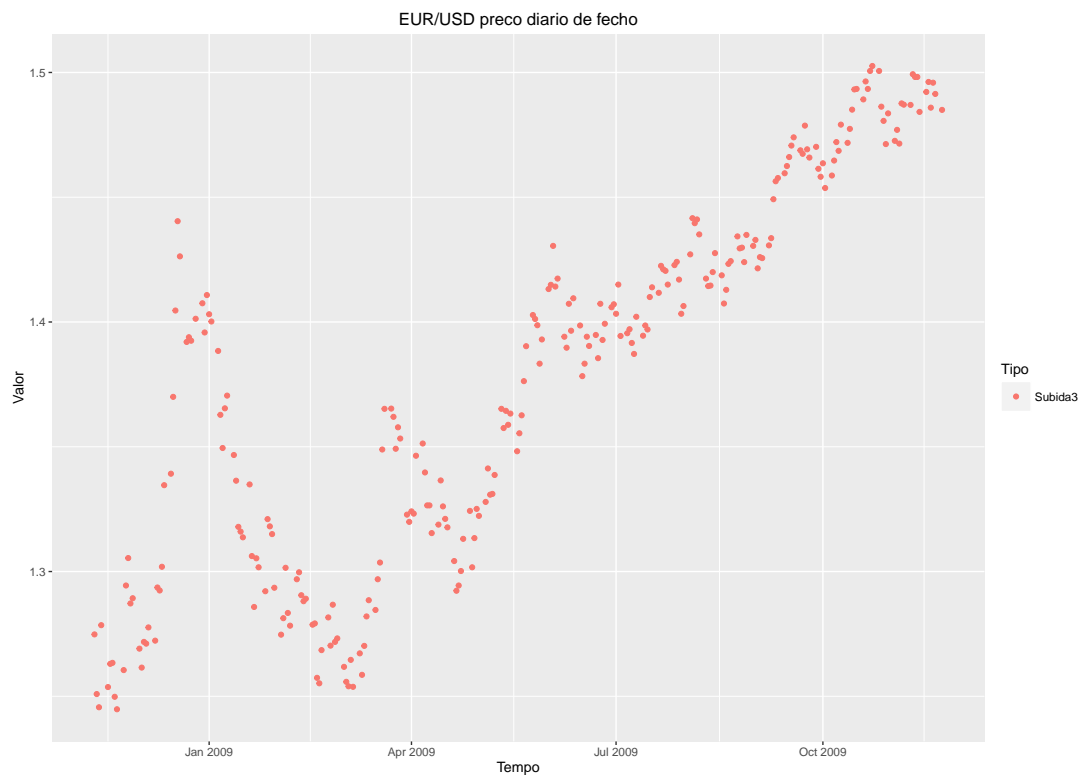
**Figura 4.7** Período 2



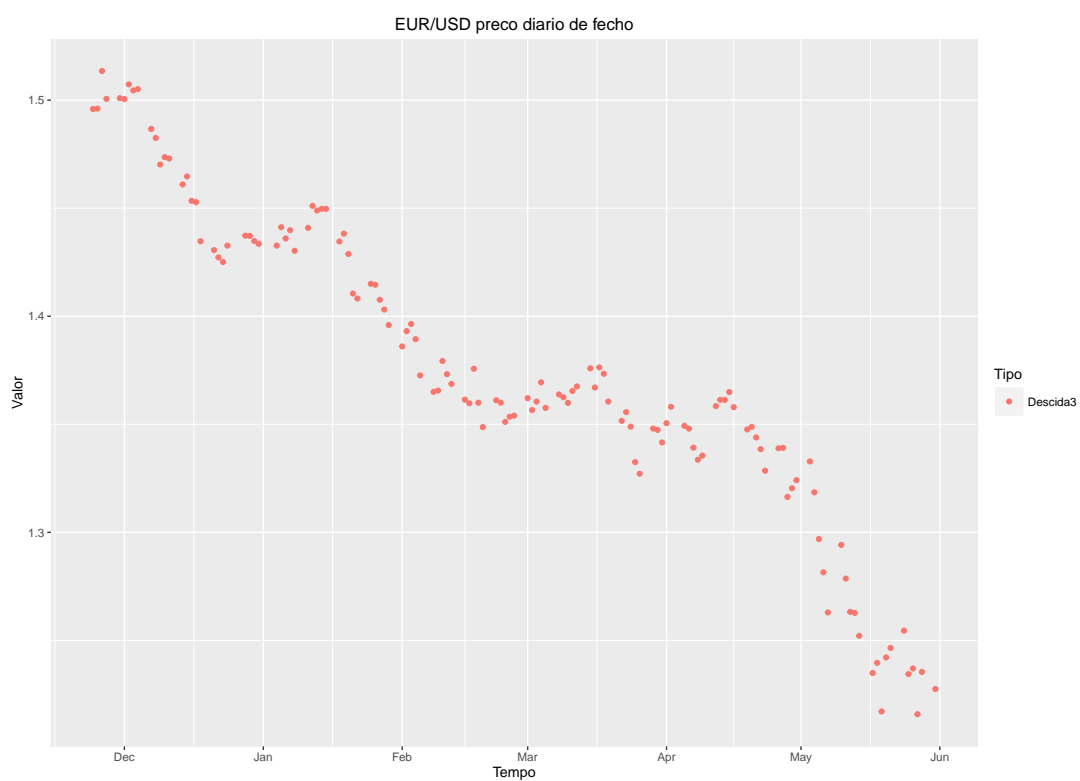
**Figura 4.8** Período 3







**Figura 4.10** Período 5



**Figura 4.11** Período 6

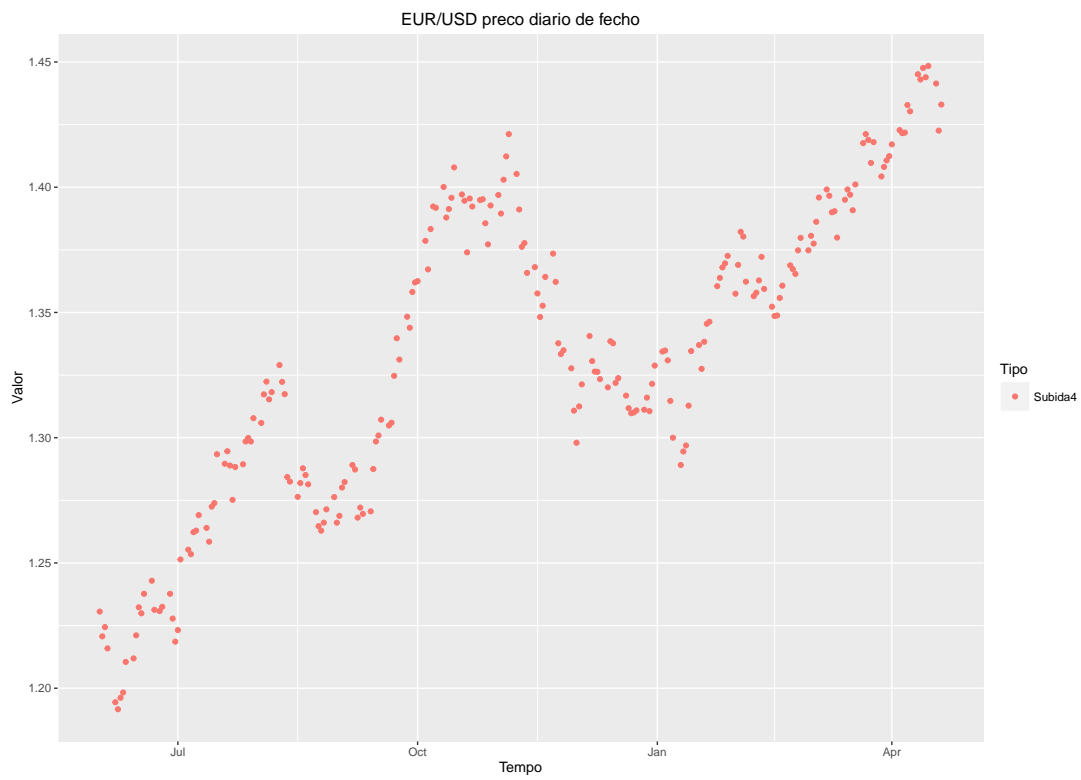
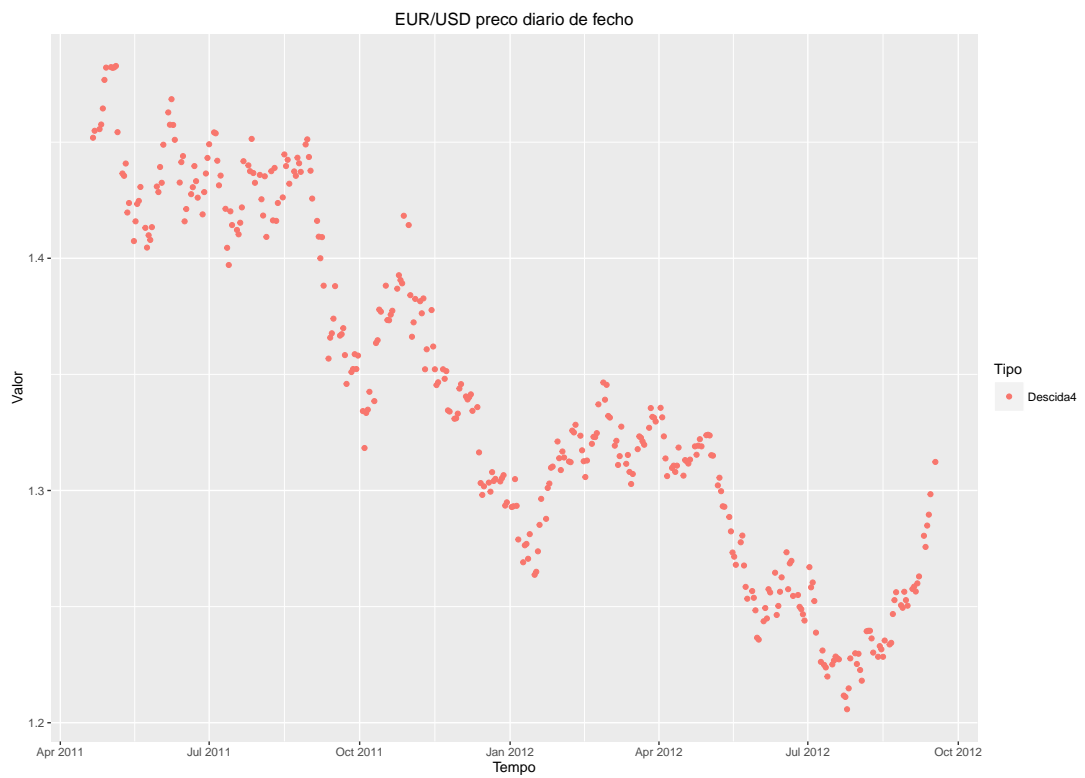
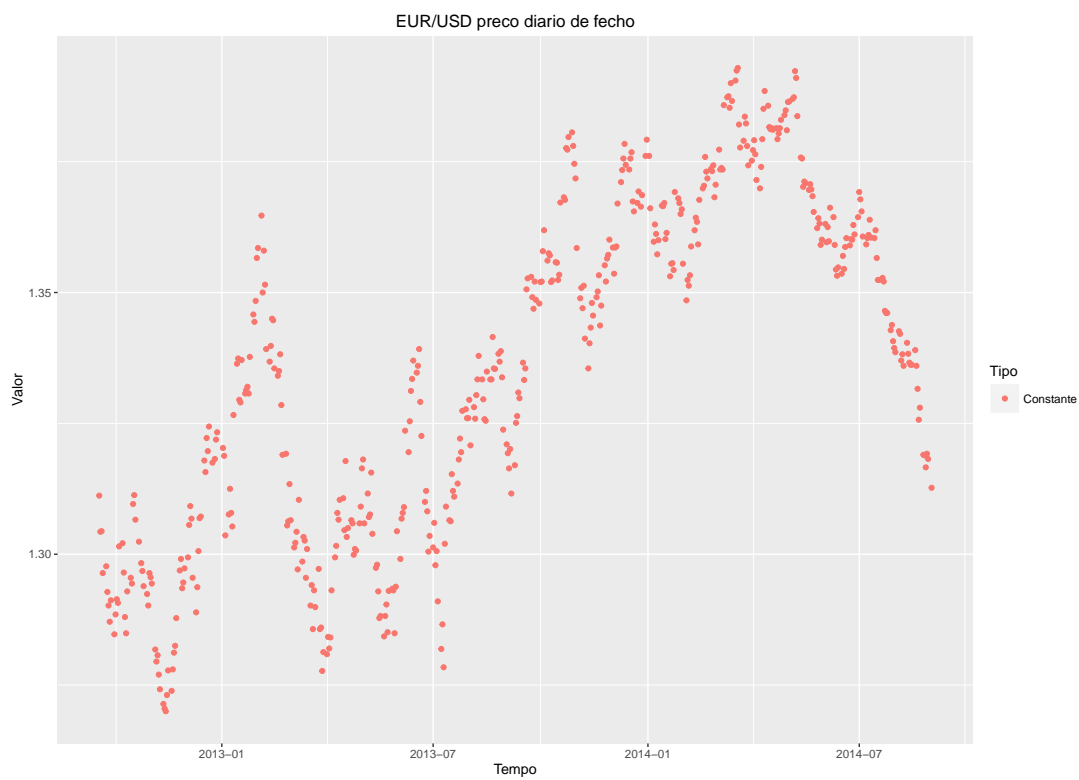


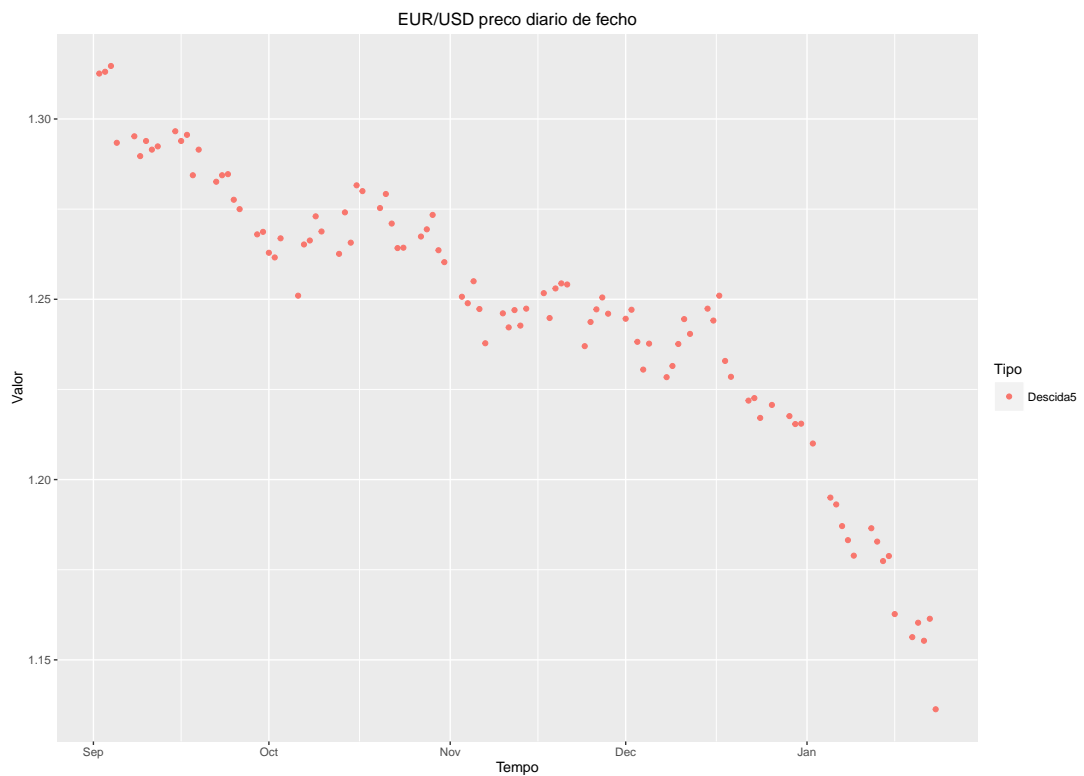
Figura 4.12 Período 7



**Figura 4.13** Período 8

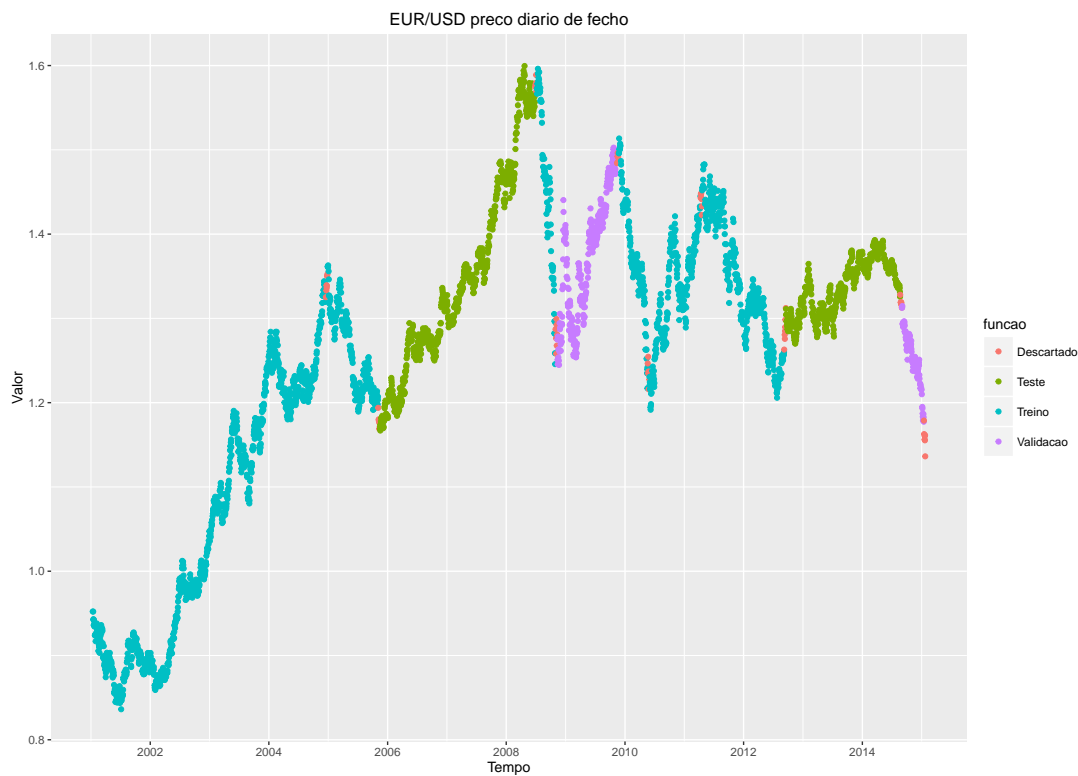


**Figura 4.14** Período 9



**Figura 4.15** Período 10

Após efetuada a divisão em 10 períodos passámos à divisão em treino, validação e teste. A forma mais natural que surgiu foi escolher aleatoriamente 6 períodos para treino, 2 para validação e 2 para teste. Visto que existe um único período constante forçámos que este faça parte do conjunto de teste, para que os modelos sejam testados num período com características diferentes do conjunto de treino. Assim, escolhemos aleatoriamente 6 dos períodos não constantes para treino, 2 para validação e o que sobrou juntamente com o período constante ficará para teste:



**Figura 4.16** Divisão em treino, validação e teste por períodos

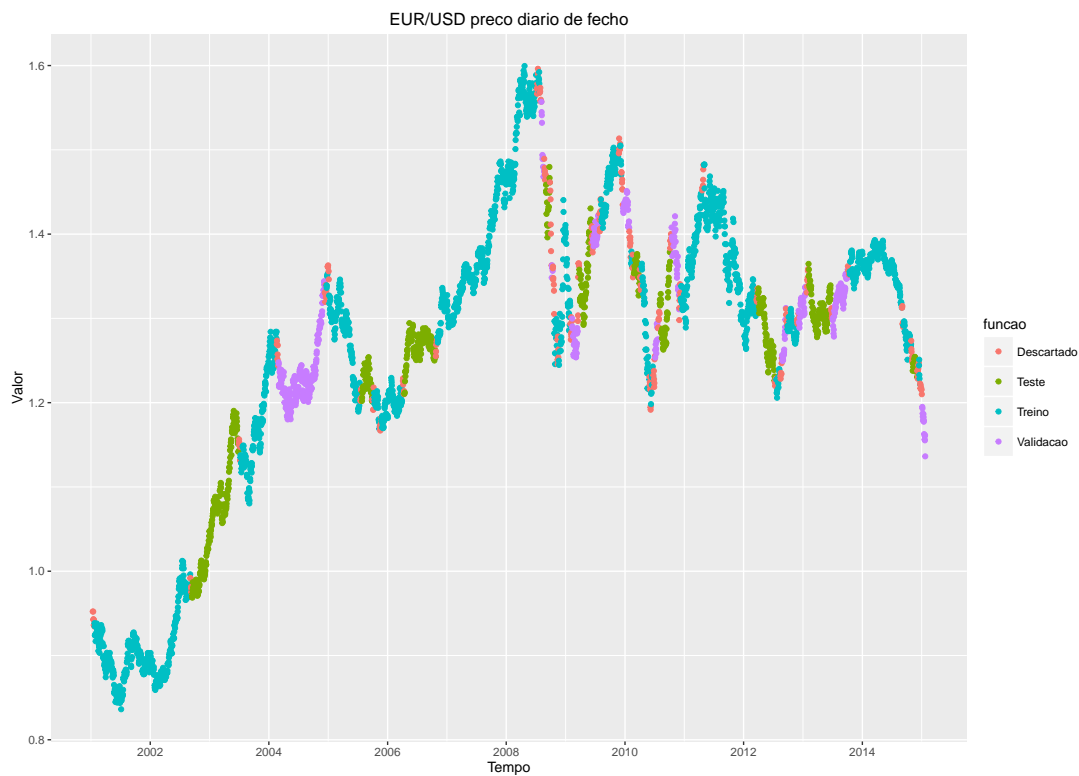
Como podemos ver na figura 4.16 com esta divisão os conjuntos de treino, validação e teste ficaram demasiado contínuos assim, decidiu-se ainda dividir cada um dos 10 períodos em treino, validação e teste aleatoriamente descartando as observações necessárias para que não haja cruzamento de informação. A divisão dentro de cada período é feita escolhendo aleatoriamente o início do teste, sabendo que este vai ter 20% das observações do período conseguimos facilmente determinar o seu fim, aplicando a mesma ideia para a validação tendo o cuidado de não haver cruzamento de informação com o teste estão fixados os dois conjuntos. Ora, as restantes observações ficam no conjunto de treino, mais uma vez tendo o cuidado de descartar observações que contenham informação sobre o teste ou a validação. Para exemplificar ver a divisão do primeiro período na figura 4.17:





**Figura 4.17** Divisão do Período 1 em treino, validação e teste

Aplicando o mesmo processo aos restantes períodos obtemos:



**Figura 4.18** Divisão em treino, validação e teste por períodos

Como podemos ver na figura 4.18 os conjuntos validação e teste estão representados ao longo de todos os períodos, em vez de estarem concentrados em apenas 2 períodos como na divisão anterior (ver figura 4.16). Esta divisão também tem uma forte componente de aleatoriedade, o que não existe na divisão por ordem cronológica (figura 4.2). Ao contrário da divisão por ordem aleatória (figura 4.3) não há qualquer tipo de cruzamento de informação entre os 3 conjuntos. Outro ponto negativo na divisão aleatória, e que também se verifica na divisão mensal (figura 4.4), é o elevado desperdício de observações.

## 4.2 Regressão logística

Este modelo desenvolvido em 1958 por David Cox [24], é um modelo linear generalizado com resposta binomial, portanto associada a dois únicos resultados, no caso deste trabalho "Up" ou "Down". Assim, a regressão logística pretende estudar a forma como a resposta varia com um conjunto de variáveis explicativas pré-específicas, nesta tese:  $Lag1, Lag2, \dots, Lag7, K, D$ .

### 4.2.1 Odds e Odds Ratio

O odds para um acontecimento corresponde ao quociente entre a probabilidade do evento acontecer e a probabilidade desse mesmo evento não acontecer:

$$\text{odds para evento} = \frac{P(\text{evento})}{1 - P(\text{evento})} \quad (4.1)$$

Um odds de 3 para um evento, por exemplo, traduz que a probabilidade de um evento acontecer é três vezes superior à probabilidade do evento não ocorrer. Um odds ratio (OR) é um quociente entre dois odds, calculados para dois grupos de observações, por exemplo:

- Seja  $Y$  a variável dicotómica que toma o valor 1 quando  $\text{Close}(t) > \text{Close}(t-1)$ , 0 caso contrário.
- Seja  $L$  a variável dicotómica que toma o valor 1 quando  $\text{Close}(t-1) > \text{Close}(t-2)$ , 0 caso contrário.

$$\text{OR} = \frac{\frac{P(Y=1|L=1)}{1-P(Y=1|L=1)}}{\frac{P(Y=1|L=0)}{1-P(Y=1|L=0)}} \quad (4.2)$$

Isto é, o quociente entre o odds para a subida nos dias em que no dia anterior se verificou uma subida e o odds para a subida nos dias em que no dia anterior se verificou uma descida. Por exemplo se  $\text{OR} = 3$ , temos que:

$$\frac{P(Y = 1|L = 1)}{P(Y = 0|L = 1)} = 3 * \frac{P(Y = 1|L = 0)}{P(Y = 0|L = 0)} \quad (4.3)$$

o que significa que o odds para a subida nos dias em que no dia anterior houve subida é três vezes superior ao odds para a subida nos dias em que no dia anterior houve uma descida.

### 4.2.2 Construção do Modelo

Regressão logística é a designação atribuída a um modelo linear generalizado com resposta binomial, portanto associada a dois únicos resultados, neste caso subida ou descida. Este modelo tenta prever a forma como a direção do EUR/USD varia com um conjunto de variáveis explicativas pré-especificadas:  $X=(Lag1,Lag2,...,Lag7,K,D)$ .

Seja  $Y$  a variável aleatória codificada por 0 quando o EUR/USD desce e 1 caso contrário. Dado um vetor de variáveis explicativas  $X=(Lag1,Lag2,...,Lag7,K,D)$ , uma observação  $x_i = (lag1, lag2, ..., lag7, k, d)$  do dia  $i$ , assume-se que:

$$Y|X = x_i \sim B(1, \pi(x_i)) \text{ ,onde} \quad (4.4)$$

$$\pi(x_i) = P(Y = 1|X = x_i) \quad (4.5)$$

é a probabilidade de o EUR/USD subir dada a observação  $x_i$ . Pelas propriedades da distribuição binomial, o valor esperado de  $Y$  para  $X=x_i$  é:

$$\mu = E(Y|X = x_i) = \pi(x_i) \quad (4.6)$$

e a variância é:

$$Var(Y|X = x_i) = \pi(x_i)(1 - \pi(x_i)) \quad (4.7)$$

Repare-se que uma modelação linear da média de  $Y|X=x_i$ , isto é de  $\pi(x)$ , conduziria a uma situação não adequada. Com efeito, fazendo:

$$\pi(x) = \beta_0 + \beta_1 Lag1 + \beta_2 Lag2 + ... + \beta_7 Lag7 + \beta_8 K + \beta_9 D \quad (4.8)$$

verificamos que o termo do lado direito é não limitado, enquanto que  $\pi(x)$  varia entre 0 e 1. Precisamos, então, de relacionar a probabilidade da resposta  $\pi(x)$  com o preditor linear

$$\eta = \beta_0 + \beta_1 Lag1 + \beta_2 Lag2 + ... + \beta_7 Lag7 + \beta_8 K + \beta_9 D \quad (4.9)$$

### 4.2.3 Modelo logit

O facto das probabilidades  $\pi_i = \pi(x_i)$  variarem entre 0 e 1 implica que os odds

$$odds_i = \frac{\pi_i}{1 - \pi_i} \quad (4.10)$$

que representam o quociente entre a probabilidade do EUR/USD não descer e a probabilidade do EUR/USD descer, variam entre 0 e  $+\infty$ . Assim, aplicando a transformação logarítmica passamos a ter:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right), \text{ que é não limitado} \quad (4.11)$$

Temos, então, que o modelo a usar é:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right) = \beta_0 + \beta_1 \text{Lag1} + \beta_2 \text{Lag2} + \dots + \beta_7 \text{Lag7} + \beta_8 K + \beta_9 D \quad (4.12)$$

Invertendo o logit tem-se que:

$$\pi = \frac{e^{\beta_0 + \beta_1 \text{Lag1} + \beta_2 \text{Lag2} + \dots + \beta_7 \text{Lag7} + \beta_8 K + \beta_9 D}}{1 + e^{\beta_0 + \beta_1 \text{Lag1} + \beta_2 \text{Lag2} + \dots + \beta_7 \text{Lag7} + \beta_8 K + \beta_9 D}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{Lag1} + \beta_2 \text{Lag2} + \dots + \beta_7 \text{Lag7} + \beta_8 K + \beta_9 D)}} \quad (4.13)$$

#### 4.2.4 Resultados

Apesar do erro ser menor que 50% os resultados não são os mais satisfatórios em nenhuma das divisões. A divisão mensal é a que apresenta o menor erro, o que pode ser explicado pelo facto de esta divisão proporcionar ao modelo um conjunto de treino bastante aleatório o que preparou melhor o modelo em relação às outras divisões.

Tipo de Divisão	Erro do Modelo
Divisão Cronológica	47.44%
Divisão Aleatória	47.80%
Divisão Mensal	45.91%
Divisão por Períodos	47.37%
Divisão por Períodos 2	46.27%

## 4.3 Árvores de Decisão

### 4.3.1 Introdução

A estrutura de uma árvore tem como início uma raiz, estando esta situada no cimo da árvore. A raiz divide-se sucessivamente através de ramos e dá origem a nós. A cada nó está associado um teste binário, que determinará o ramo a seguir no próximo passo. Caso um nó não tenha ramificações, chama-se folha ou nó terminal e determinará a classe a que determinado objecto pertence. Este tipo de modelo apareceu nos anos 60 e propõe uma estrutura hierárquica que se parece com uma árvore. A metodologia base é "dividir para conquistar"[25] e atualmente há diferentes algoritmos para a construção da estrutura. Nesta tese foi usado o algoritmo CART ("Classification And Regression Trees") proposto por Breiman em 1984 [26].

### 4.3.2 Poda

Independentemente do algoritmo que se escolha, a árvore obtida com base na amostra de treino tem o nome de árvore máxima. Geralmente, esta árvore máxima está demasiado ajustada ("overfitting") ao conjunto de treino o que resulta em resultados pouco satisfatórios quando aplicada noutros conjuntos. Assim, para combater este problema, geralmente poda-se a árvore máxima, isto é removem-se nós. Existem essencialmente dois métodos: [27]

- **métodos pré-poda:** consiste em limitar o crescimento da árvore, para isso é fixado um critério de paragem do algoritmo de construção da árvore.
- **métodos pós-poda:** deixar o algoritmo criar a árvore máxima e posteriormente podá-la, originando, assim, uma subárvore.

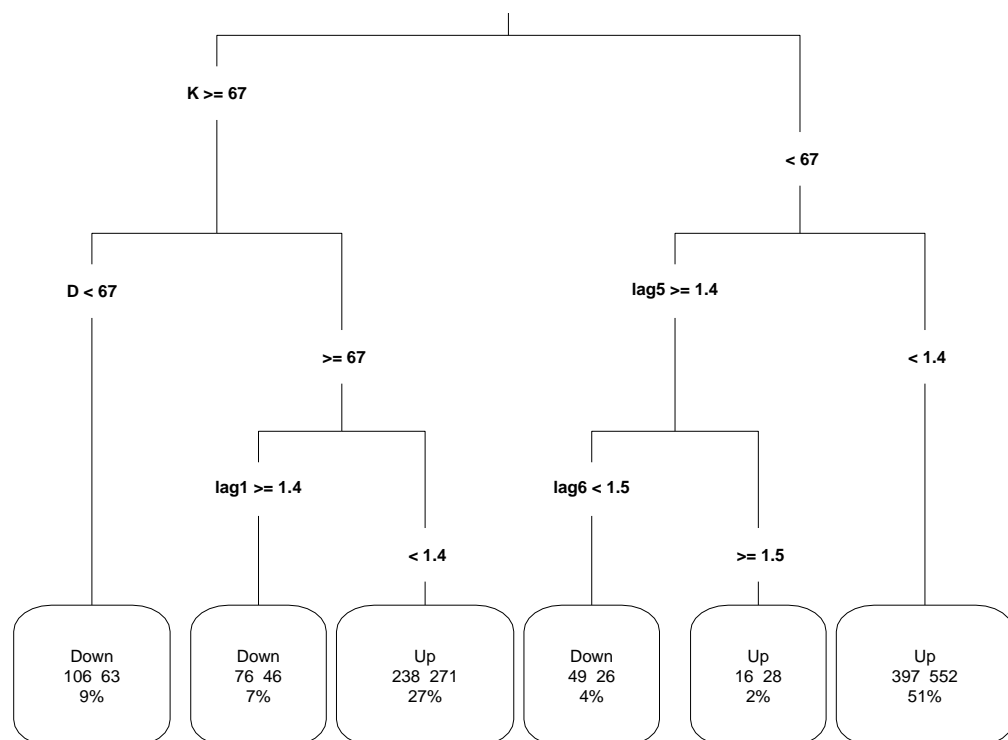
O método mais popular na comunidade académica é o método pós-poda e é o adoptado nesta pesquisa. Mais especificamente, foi usada a técnica proposta por Breiman em 1984 [26]: poda por minimização do custo-complexidade, que consiste na eliminação de ramos da árvore máxima.



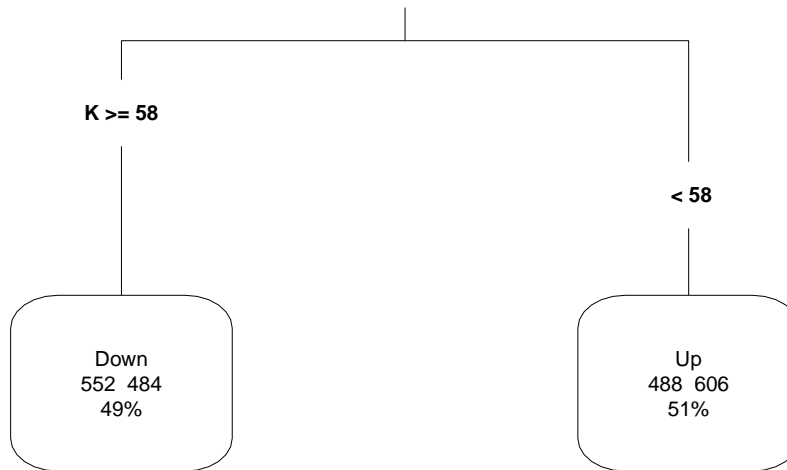
### 4.3.3 Resultados

Tipo de Divisão	Erro do Modelo
Divisão Cronológica	48.27%
Divisão Aleatória	48.21%
Divisão Mensal	45.71%
Divisão por Períodos	48.22%
Divisão por Períodos 2	48.76%

Os resultados são de novo pouco satisfatórios, tendo, até piorado comparativamente à regressão logística. Também é curioso a diferença da complexidade da estrutura da árvore da divisão aleatória e por períodos 2 (ver figura 4.19 e 4.20 respectivamente.)



**Figura 4.19** Árvore obtida usando a divisão de dados aleatória



**Figura 4.20** Árvore obtida usando a divisão de dados por períodos 2

## 4.4 Análise Discriminante

### 4.4.1 Linear

#### Introdução

A Análise Discriminante Linear (LDA), também conhecida por modelo normal homocedástico é geralmente usada para reduzir o número de variáveis em problemas de classificação. O objectivo é projectar as observações num espaço de menor dimensão com uma boa separação de classes de forma a reduzir o custo computacional e problemas de "overfitting". Este modelo foi formulado em 1936 por Ronald A. Fisher [28] originalmente para problemas de duas classes. Posteriormente (em 1948) foi generalizado para um número finito de classes por C. R. Rao [29]. Porém, visto que o nosso problema tem apenas duas classes vamos usar

a versão original. É de realçar que este método assume homoscedasticidade.

### Resultados

À semelhança do que aconteceu nos últimos dois modelos a divisão mensal voltou a ser a divisão com o menor erro.

Tipo de Divisão	Erro do Modelo
Divisão Cronológica	47.30%
Divisão Aleatória	47.80%
Divisão Mensal	45.91%
Divisão por Períodos	47.29%
Divisão por Períodos 2	46.27%

#### 4.4.2 Quadrática

##### Introdução

Visto que na realidade muitas vezes não se pode assumir homoscedasticidade, utiliza-se o modelo normal heterocedástico ou análise discriminante quadrática (QDA) que basicamente é a generalização da análise discriminante linear para situações em que não se pode assumir homoscedasticidade. Normalmente utiliza-se QDA em vez de LDA quando:

1. Assumir homoscedasticidade não é viável.
2. O conjunto de treino é grande. Por outro lado, quando o conjunto de treino é pequeno geralmente o modelo QDA cria problemas de "overfitting".

**Resultados**

Tipo de Divisão	Erro do Modelo
Divisão Cronológica	48.83%
Divisão Aleatória	48.21%
Divisão Mensal	50.50%
Divisão por Períodos	47.37%
Divisão por Períodos 2	47.79%

Ao contrário do que aconteceu com os outros modelos, a divisão mensal foi a que apresentou o maior erro, o que pode estar relacionado com o facto de nesta divisão haver uma grande remoção de observações para que não haja cruzamento de informação entre os conjuntos. O desperdício de observações levou a que ficassemos com um número de observações relativamente pequeno o que pode ter causado "overfitting" do modelo ao conjunto do treino.



## 5 Modelos aplicados a dados de alta frequência

### 5.1 Plataforma MetaTrader 4

Lançado em 2005, O MetaTrader 4, também conhecido como MT4, é uma plataforma de “trading” (compra e venda de ativos financeiros) muito popular. Desenvolvido pela MetaQuotes Software, este software é constituído por uma componente cliente e uma componente servidor. A componente servidor é assegurada pela corretora que posteriormente disponibiliza a componente cliente aos seus clientes, que a utilizam para ter acesso a preços do mercado em tempo real, colocar ordens entre outras coisas. Todas as estratégias apresentadas neste capítulo foram implementadas em ambiente MetaTrader 4 aproveitando a linguagem de programação do próprio software: MetaQuotes Language 4 (MQL4). A cada ordem que o software coloca no mercado estão associados vários parâmetros, dos quais se destacam:

- Tipo de ordem (compra ou venda)
- Volume
- Preço
- Stoploss (pior cenário que estamos dispostos a correr, mal o ativo atinja o valor indicado a ordem será fechada)
- Takeprofit (objectivo de ganho da transação, mal o ativo atinja este valor a ordem será fechada)

As estratégias foram testadas em diversas corretoras sempre com um "spread" (valor absoluto da diferença entre o preço de compra (Ask) e de preço venda (Bid) num dado instante) constante de 0.8 pips (1 pips = 0.0001).

### 5.1.1 Estratégia 1

Desenhada para ser implementada em qualquer par cambial, esta estratégia tem como base cinco valores de referencia: dois alvos de subida ( $s1(t)$  e  $s2(t)$ ), dois alvos de descida ( $d1(t)$  e  $d2(t)$ ) e um valor base ( $b(t)$ ), a cada instante o algoritmo verifica se o preço da mercadoria é superior ou inferior ao valor base, caso seja superior abre duas ordens de compra cada uma delas com o Takeprofit num dos alvos de subida e com o StopLoss no maior alvo de descida; caso seja inferior ao valor base abre duas ordens de vende cada uma delas com o Takeprofit nos respectivos alvos de descida e o Takeprofit no menor alvo de subida.

$$b(t) = \frac{high(t) + low(t) + close(t)}{3} \quad (5.1)$$

$$d1(t) = 2 * b(t) - high(t) \quad (5.2)$$

$$d2(t) = b(t) - 2 * (high(t) - low(t)) \quad (5.3)$$

$$s1(t) = 2 * b(t) - low(t) \quad (5.4)$$

$$s2(t) = b(t) + 2 * (high(t) - low(t)) \quad (5.5)$$

É óbvio que:

$$high(t) \geq b(t) \geq low(t) \quad (5.6)$$

Vamos, então, provar que

$$s2(t) \geq s1(t) \geq d2(t) \geq d1(t) \quad (5.7)$$

Vamos começar por provar que  $s2(t) \geq s1(t)$ :

$$\begin{aligned}
 s2(t) \geq s1(t) &\iff s2(t) - s1(t) \geq 0 \iff \\
 &\iff b(t) + 2 * (high(t) - low(t)) - (2 * b(t) - low(t)) \geq 0 \iff \\
 &\iff b(t) + 2 * high(t) - 2 * low(t) - 2 * b(t) + low(t) \geq 0 \iff \\
 &\iff 2 * high(t) - b(t) - low(t) \geq 0, \text{ o que é sempre verdade por (5.6)}
 \end{aligned} \tag{5.8}$$

Vamos, agora, provar que  $s1(t) \geq d1(t)$ :

$$\begin{aligned}
 s1(t) \geq d1(t) &\iff s1(t) - d1(t) \geq 0 \iff \\
 &\iff 2 * b(t) - low(t) - (2 * b(t) - high(t)) \geq 0 \iff \\
 &\iff 2 * b(t) - low(t) - 2 * b(t) + high(t) \geq 0 \iff \\
 &\iff high(t) - low(t) \geq 0, \text{ o que é sempre verdade por (5.6)}
 \end{aligned} \tag{5.9}$$

Resta-nos verificar que  $d1(t) \geq d2(t)$ :

$$\begin{aligned}
 d1(t) \geq d2(t) &\iff d1(t) - d2(t) \geq 0 \iff \\
 &\iff 2 * b(t) - high(t) - (b(t) - 2 * (high(t) - low(t))) \geq 0 \iff \\
 &\iff 2 * b(t) - high(t) - b(t) + 2 * high(t) - 2 * low(t) \geq 0 \iff \\
 &\iff b(t) + high(t) - 2 * low(t) \geq 0, \text{ o que é sempre verdade por (5.6)}
 \end{aligned} \tag{5.10}$$



### 5.1.2 Estratégia 2

Construída para ser aplicada no par EUR/USD, esta estratégia tem por base duas curvas do oscilador estocástico com diferentes períodos  $k1$ . Ao contrário da estratégia 1, esta apenas coloca ordens longas no mercado. O sinal de compra é dado pela a intersecção das curvas  $D(t)$  acima do valor 50:

$$Takeprofit = Ask + 600 \text{ pips} \quad (5.11)$$

$$Stoploss = Ask - 3000 \text{ pips} \quad (5.12)$$

As ordens serão fechadas caso o preço atinja o Takeprofit ou Stoploss ou haja um sinal de venda, isto é que as curvas  $D(t)$  se intersectem abaixo do valor 50.

Para testar a estratégia foram feitas simulações usando o histórico do EUR/USD do MT4 de 2014 e 2015, os resultados mostraram que o algoritmo ia ganhando dinheiro lentamente e que num dia “mau” perdia tudo o que tinha ganho nos dias anteriores (e possivelmente mais ainda). Para tentar “fugir” a estes momentos do mercado, o algoritmo foi programado para fazer uma auto-avaliação do desempenho da estratégia em tempo real fazendo uma contagem do número consecutivo de ordens com retorno positivo e negativo, assim quando o número de consecutivo de ordens com retorno negativo atinge os 25 a estratégia faz uma pausa (não coloca ordens) de 12 horas, após este período volta ao ativo, se o mau desempenho se continuar a verificar e atingir as 75 ordens sempre com retorno negativo é feita nova pausa desta vez de 24 horas. Por outro lado, caso o número de ordens com retorno positivo consecutivas atinja os 25 o volume das ordens futuras duplica até que haja uma ordem com retorno negativo. Os resultados melhoraram bastante com a introdução da auto-avaliação em tempo real, tendo o algoritmo uma rentabilidade YTD (01/01/2016 a 31/03/2016) de 20%.

### 5.1.3 Estratégia 3

Desenhada para ser implementada no par EUR/USD, esta estratégia tenta captar a tendência do mercado cambial tomando uma posição longa (ordem de compra) sempre que estiverem reunidas as seguintes condições:

1.  $Ask > EMA(price(t), 15)$
2.  $MACD-Signal < 0$  e  $MACD-Main < 0$
3.  $ATR(t) > 25$

A ordem terá associado um Takeprofit e um Stoploss com as seguintes fórmulas:

$$Takeprofit = Ask + 4 * ATR(t) \quad (5.13)$$

$$Stoploss = Ask - 2 * ATR(t) \quad (5.14)$$

A posição será fechada caso um dos três acontecimentos se realize:

1. Acontecimento 1:  
 $Ask \geq Takeprofit$
2. Acontecimento 2:  
 $Ask \leq Stoploss$
3. Acontecimento 3:
  - a)  $Ask < EMA(Ask, 15)$
  - b)  $MACD-Signal > 0$  e  $MACD-Main > 0$
  - c)  $ATR(t) > 25$

Caso se verifiquem:

1.  $Bid < EMA(Bid(t), 15)$

2.  $MACD-Signal > 0$  e  $MACD-Main > 0$

3.  $ATR(t) > 25$

é aberta uma ordem de venda com os seguintes valores associados:

$$Takeprofit = Bid - 4 * ATR(t) \quad (5.15)$$

$$Stoploss = Bid + 2 * ATR(t) \quad (5.16)$$

A ordem será fechada caso o preço atinja o Takeprofit ou o Stoploss.

## 6 Conclusão

Este trabalho teve como principal objectivo a construção de estratégias aplicados a dados de alta frequência de forma a atingir os objectivos do estágio proposto pelo Banco L. J. Carregosa. Adicionalmente, foi feita uma pesquisa sobre as diferentes formas de dividir o conjunto de dados em treino, validação e teste usadas dentro da comunidade académica e económica, assim como foram propostas novas formas de o fazer. Para aplicar as diferentes divisões de dados, foram construídos modelos preditivos, baseados em análise discriminante linear, análise discriminante quadrática, regressão logística e árvores de decisão com o intuito de prever a direcção do preço de fecho diário do par EUR/USD.

Os resultados deste trabalho corroboram a grande dificuldade de prever o mercado financeiro, mais especificamente o mercado cambial. Porém, surpreendentemente a segunda estratégia das estratégias aplicadas a dados de alta frequência teve resultados bastante positivos, podendo mesmo vir a ser usada pelo Banco L. J. Carregosa para gestão interna de carteiras ou disponibilizada aos seus investidores para uso próprio. Para além da criação das estratégias, o estágio teve um impacto directo na implementação da plataforma MetaTrader 4 no banco. As novas divisões de dados propostas nesta tese, apresentaram-se consistentes e podem vir ser utilizadas noutros trabalhos no futuro. Relativamente aos modelos aplicados a dados de baixa frequência, apesar de terem erros inferiores a 50%, terá que ser feita uma revisão cuidada dos custos de mercado antes de investir nestes modelos.



# Bibliografia

- [1] B. F. I. Settlements, “Foreign exchange turnover in april 2013: preliminary global results,” September 2013. [Online]. Available: <http://www.bis.org/publ/rpfx13fx.pdf>
- [2] M. P. Taylor and H. Allen, “The use of technical analysis in the foreign exchange market,” *Journal of international Money and Finance*, vol. 11, no. 3, pp. 304–314, 1992.
- [3] B. LeBaron, “Chaos and nonlinear forecastability in economics and finance,” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 348, no. 1688, pp. 397–404, 1994.
- [4] G. F. Luger, *Artificial intelligence: structures and strategies for complex problem solving*. Pearson education, 2005.
- [5] C. W. Smith Jr, C. W. Smithson, and L. M. Wakeman, “The market for interest rate swaps,” *Financial Management*, pp. 34–44, 1988.
- [6] R. Barroso, “Os argumentos do estado e do santander no caso dos "swaps",” *Jornal de Negocios*, 2016. [Online]. Available: <http://www.jornaldenegocios.pt/mercados/taxas-de-juro/detalhe/os-argumentos-do-estado-e-do-santander-no-caso-dos-swaps>
- [7] M. Burton, R. F. Nesiba, and B. Brown, *An introduction to financial markets and institutions*. Routledge, 2015.
- [8] C. Bastardo, “Gestão de activos financeiros-back to basis,” *Lisboa: Escolar Editora. BCP (2010), Disponivel em*, vol. 4, no. 12, p. 2010, 2011.
- [9] I. O. for Standardization. Currency codes - iso 4217. [Online]. Available: [http://www.iso.org/iso/home/standards/currency\\_codes.htm](http://www.iso.org/iso/home/standards/currency_codes.htm)

- [10] J. Regnault, *Calcul des chances et philosophie de la bourse*. Mallet-Bachelier, 1863.
- [11] L. Bachelier, *Théorie de la spéculation*. Gauthier-Villars, 1900.
- [12] P. H. Cootner, "The random character of stock market prices," 1964.
- [13] E. F. Fama, "Random walks in stock market prices," *Financial analysts journal*, vol. 51, no. 1, pp. 75–80, 1995.
- [14] P. Tenti, "Forecasting foreign exchange rates using recurrent neural networks," *Applied Artificial Intelligence*, vol. 10, no. 6, pp. 567–582, 1996.
- [15] A. Serletis, A. G. Malliaris, M. J. Hinich, and P. Gogas, "Episodic nonlinearity in leading global currencies," *Open Economies Review*, vol. 23, no. 2, pp. 337–357, 2012.
- [16] M. Alamili, "Exchange rate prediction using support vector machines," Ph.D. dissertation, DELFT UNIVERSITY OF TECHNOLOGY, 2011.
- [17] R. Meese and K. Rogoff, "The out-of-sample failure of empirical exchange rate models: sampling error or misspecification?" in *Exchange rates and international macroeconomics*. University of Chicago Press, 1983, pp. 67–112.
- [18] Y.-W. Cheung, M. D. Chinn, and A. G. Pascual, "Empirical exchange rate models of the nineties: Are any fit to survive?" *Journal of International Money and finance*, vol. 24, no. 7, pp. 1150–1175, 2005.
- [19] J. B. Maverick. What is the exponential moving average (ema) formula and how is the ema calculated? [Online]. Available: <http://www.investopedia.com/ask/answers/122314/what-exponential-moving-average-ema-formula-and-how-ema-calculated.asp>
- [20] J. J. Murphy, *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin, 1999.
- [21] J. W. Wilder, *New concepts in technical trading systems*. Trend Research, 1978.

- [22] B. L.P., *Definition of Average True Rate*.
- [23] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics Springer, Berlin, 2001, vol. 1.
- [24] D. R. Cox, “The regression analysis of binary sequences,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 215–242, 1958.
- [25] A. Januário, “Exchange rate prediction using support vector machines,” Ph.D. dissertation, Faculdade de Ciências da Universidade do Porto, 2012.
- [26] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*. CRC press, 1984.
- [27] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.
- [28] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [29] C. R. Rao, “The utilization of multiple measurements in problems of biological classification,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 10, no. 2, pp. 159–203, 1948.